Regularized estimation of the nominal response model

Michela Battauz

Department of Economics and Statistics
University of Udine (Italy)

July 10, 2019
The nominal response model (Bock, 1972)

- Probability of giving response 
  \( k = 0, \ldots, m_j - 1 \) to item \( j \)

\[
P(Y_{ij} = k|\theta_i) = \frac{e^{\alpha_{jk}\theta_i + \beta_{jk}}}{\sum_{h=0}^{m_j-1} e^{\alpha_{jh}\theta_i + \beta_{jh}}},
\]

where

- \( \theta_i \) is the latent variable of subject \( i \),
- \( \alpha_{jk} \) are slope parameters,
- \( \beta_{jk} \) are intercept parameters.

- For identifiability, \( \alpha_{j0} = 0 \) and \( \beta_{j0} = 0 \) \( \forall j \).
- It is the most \textbf{flexible} IRT model for polytomous responses.
- However, it involves the estimation of \textbf{many parameters}. 

![Graph showing probability distribution for the nominal response model](image-url)
 Usually, **marginal likelihood method**, which requires the maximization of the marginal log-likelihood function

\[
\ell(\alpha, \beta) = \sum_{j=1}^{J} \log \int_{\mathbb{R}} \prod_{k=0}^{m_j-1} P(Y_{ij} = k | \theta_i) I(Y_{ij} = k) \phi(\theta_i) d\theta_i
\]

where

- \( J \) is the number of items,
- \( \alpha \) is a vector containing all the slope parameters,
- \( \beta \) is a vector containing all the intercept parameters,
- \( I(\cdot) \) is the indicator function,
- \( \phi(\cdot) \) denotes the density of the standard normal variable.
The lasso method (Tibshirani, 1996)

- First proposed for the linear regression model.
- A constraint is added to the least square problem with the effect of shrinking some coefficients and setting others to zero.
- Corresponds to the minimization of a loss function with a penalty term.

$$\min_{\beta} \left\{ \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij} \beta_j)^2 \right\} + \lambda \sum_{j=1}^{p} |\beta_j|$$
If $\alpha_{jk} = \alpha_{jh} \Rightarrow$ categories $k$ and $h$ can be **collapsed** (Thissen and Cai, 2016).

**Proposal:** penalty that encourages the slope parameters of the same item to assume the same value. The penalized log-likelihood function:

$$
\ell_p(\alpha, \beta) = \ell(\alpha, \beta) - \lambda \sum_{j=1}^{J} \sum_{k=0}^{m_j-2} \sum_{h=k+1}^{m_j-1} |\alpha_{jk} - \alpha_{jh}|.
$$

Similar to fused lasso (Tibshirani et al., 2005) but here there is not a natural order of the slope coefficients.

Since $\alpha_{j0} = 0 \ \forall j$, the penalty constrains the slope parameters toward zero: for $\lambda \to \infty$, $\alpha_{jk} = 0 \ \forall j, k$. 
Probability curves for increasing values of $\lambda$

\[ \lambda = 0 \]

\begin{align*}
\text{prob} & \quad \theta \\
-4 & \quad -2 & \quad 0 & \quad 2 & \quad 4
\end{align*}

\begin{align*}
\lambda & \quad \theta & \quad \text{prob} \\
0 & \quad 0 & \quad 0.0 \\
1 & \quad 1 & \quad 0.5 \\
2 & \quad 2 & \quad 1.0 \\
3 & \quad 3 & \quad 1.5 \\
4 & \quad 4 & \quad 2.0 \\
5 & \quad 5 & \quad 2.5 \\
6 & \quad 6 & \quad 3.0
\end{align*}

---

Michela Battauz

Regularized estimation of the nominal model

July 10, 2019 6 / 20
Probability curves for increasing values of $\lambda$
Probability curves for increasing values of $\lambda$

\[ \lambda = 0.2 \]

\[ \theta \]

\[ \text{prob} \]

\[ \text{slope} \quad \text{intercept} \]
Probability curves for increasing values of $\lambda$

$\lambda = 0.3$

---

Michela Battauz
Regularized estimation of the nominal model
July 10, 2019
Probability curves for increasing values of $\lambda$

$\lambda = 0.4$

- $\theta$
- $\text{prob}$

- $\lambda$
- $\text{slope}$
- $\text{intercept}$
Probability curves for increasing values of $\lambda$

$\lambda = 0.5$

- $\theta$ vs prob
- $\lambda$ vs slope and intercept
Probability curves for increasing values of $\lambda$

$\lambda = 0.6$

- $\theta$ vs prob
- $\lambda$ vs slope and intercept
Probability curves for increasing values of $\lambda$

$\lambda = 0.7$

The graphs show probability curves for increasing values of $\lambda$. The x-axis represents $\theta$, and the y-axis represents the probability. Different curves are plotted for varying values of $\lambda$. The slopes and intercepts are indicated for reference.
Probability curves for increasing values of $\lambda$

$\lambda = 0.8$

- $\theta$
  - prob
  - slope
  - intercept

Michela Battauz
Regularized estimation of the nominal model
July 10, 2019

6 / 20
Probability curves for increasing values of $\lambda$

$\lambda = 0.9$

- $\theta$
- $\text{prob}$

- $\lambda$
- $\text{slope}$
- $\text{intercept}$
Probability curves for increasing values of $\lambda$
Probability curves for increasing values of $\lambda$

$\lambda = 1.1$

---

Michela Battaaz
Regularized estimation of the nominal model
July 10, 2019
Probability curves for increasing values of $\lambda$
Probability curves for increasing values of $\lambda$
Probability curves for increasing values of $\lambda$

$\lambda = 1.4$

---

Michela Battauz  
Regularized estimation of the nominal model  
July 10, 2019  
6 / 20
Probability curves for increasing values of $\lambda$
Probability curves for increasing values of $\lambda$

$\lambda = 1.6$

---

Michela Battauz
Regularized estimation of the nominal model
July 10, 2019
Probability curves for increasing values of $\lambda$
Probability curves for increasing values of $\lambda$

$\lambda = 1.8$

---

Michela Battauz

Regularized estimation of the nominal model

July 10, 2019  6 / 20
Probability curves for increasing values of $\lambda$

$\lambda = 1.9$

---

Michela Battauz
Regularized estimation of the nominal model
July 10, 2019 6 / 20
Probability curves for increasing values of $\lambda$

$\lambda = 2$

---

Michela Battauz
Regularized estimation of the nominal model
July 10, 2019
Probability curves for increasing values of $\lambda$

$\lambda = 2.1$

---

slope

intercept
Probability curves for increasing values of $\lambda$

$\lambda = 2.2$

---

Michela Battauz
Regularized estimation of the nominal model
July 10, 2019
6 / 20
Probability curves for increasing values of $\lambda$
Probability curves for increasing values of $\lambda$

$\lambda = 2.4$

slope  intercept
Probability curves for increasing values of $\lambda$

\[\lambda = 2.5\]

\[\theta\]

\[\text{prob}\]

\[\text{slope}\]

\[\text{intercept}\]

Michela Battauz

Regularized estimation of the nominal model

July 10, 2019
Probability curves for increasing values of $\lambda$

$\lambda = 2.6$

---

Michela Bhattauz  
Regularized estimation of the nominal model  
July 10, 2019
Probability curves for increasing values of $\lambda$
Probability curves for increasing values of $\lambda$
Probability curves for increasing values of $\lambda$

$$\lambda = 2.9$$

---

**Michela Battauz**
Regularized estimation of the nominal model
July 10, 2019

---
Probability curves for increasing values of $\lambda$

$\lambda = 3$

$\theta$

$\lambda$

$\text{prob}$

slope

intercept

---

Michela Battauz

Regularized estimation of the nominal model

July 10, 2019
Probability curves for increasing values of $\lambda$

$\lambda = 3.1$

The graph shows probability curves for increasing values of $\lambda$. The x-axis represents $\theta$, and the y-axis represents the probability prob. The curves are labeled as slope and intercept.
Probability curves for increasing values of $\lambda$

$\lambda = 3.2$

The graphs illustrate the probability curves for increasing values of $\lambda$. The left graph shows the relationship between $\theta$ and prob, while the right graph illustrates the relationship between $\lambda$ and prob, with dashed lines indicating the intercept and solid lines indicating the slope.
Probability curves for increasing values of $\lambda$

\[ \lambda = 3.3 \]

- The graphs show the probability curves for different values of $\lambda$.
- The x-axis represents $\theta$ and the y-axis represents prob.
- The graphs include lines indicating slope and intercept.

Michela Battaui

Regularized estimation of the nominal model

July 10, 2019
probability curves for increasing values of $\lambda$
Probability curves for increasing values of $\lambda$

$\lambda = 3.5$

- $\theta$
- $\lambda$
- $\text{prob}$
- $\text{slope}$
- $\text{intercept}$

Michela Battaux
Regularized estimation of the nominal model
July 10, 2019
Probability curves for increasing values of $\lambda$

$\lambda = 3.6$

---

Michela Battauz  
Regularized estimation of the nominal model  
July 10, 2019
Probability curves for increasing values of $\lambda$

$\lambda = 3.7$

---

Michela Battauz
Regularized estimation of the nominal model
July 10, 2019
Probability curves for increasing values of $\lambda$
Probability curves for increasing values of $\lambda$
Probability curves for increasing values of $\lambda$

$\lambda = 4$

---

Michela Battauz
Regularized estimation of the nominal model
July 10, 2019 6 / 20
Probability curves for increasing values of $\lambda$
Probability curves for increasing values of $\lambda$

![Graph showing probability curves for different values of $\lambda$.]
Probability curves for increasing values of $\lambda$

$\lambda = 4.3$

---

Michela Battauz  
Regularized estimation of the nominal model  
July 10, 2019  6 / 20
Probability curves for increasing values of $\lambda$
Probability curves for increasing values of $\lambda$

\[ \lambda = 4.5 \]

\[ \theta \]

\[ \text{prob} \]

\[ 0, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 \]

\[ -4, -2, 0, 2, 4 \]

\[ \lambda = 4.5 \]

\[ \theta \]

\[ \text{prob} \]

\[ 0, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 \]

\[ -4, -2, 0, 2, 4 \]

\[ \lambda \]

\[ 0, 1, 2, 3, 4, 5 \]

\[ 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0 \]

---

Michela Battauz

Regularized estimation of the nominal model

July 10, 2019
Probability curves for increasing values of $\lambda$

![Graph showing probability curves for increasing values of $\lambda$.]
Probability curves for increasing values of $\lambda$
Probability curves for increasing values of $\lambda$

\[
\lambda = 4.9
\]
Probability curves for increasing values of $\lambda$
Adaptive version (Zou, 2006) of the penalty

\[ \ell_p(\alpha, \beta) = \ell(\alpha, \beta) - \lambda \sum_j \sum_{m_j-2}^{m_j-1} \sum_{k=0} h=k+1 |\alpha_{jk} - \alpha_{jh}| w_{jkh}, \]

where \( \hat{\alpha}^{MLE}_{jk} \) denotes the maximum likelihood estimate of the slope parameters.
The maximization of the penalized log-likelihood function not simple because it is **not differentiable everywhere**.

Explored various algorithms:
- alternating direction method of multipliers (Hastie at al., 2015),
- proximal gradient (Hastie at al., 2015),
- approximation of the absolute value \( |\xi| \approx \sqrt{\xi^2 + c} \) (Tutz and Gertheiss, 2014).

The tuning parameter \( \lambda \) is selected using **cross-validation**.

R (and C++) used for all analyses.
Year 2011, Mathematics, 8th grade.

- Items in Block M01: 5 multiple choice and 6 constructed response questions $\Rightarrow 52$ parameters.
- Country: United States $\Rightarrow 731$ subjects.
## Scoring guide of item M032761

| **Correct Response** |  
|----------------------|---|
| **20**               | Both expressions correct in simplified form  
|                      | Red tiles: $4(n - 1); 4n - 4$; or correct verbal expression  
|                      | Total tiles: $n^2; n \times n$; or correct verbal expression, such as “square the number” or “multiply by itself”  
| **21**               | Both expressions correct with expression for red tiles in the form of total number of tiles minus number of black tiles e.g., $n^2 - (n - 2)^2$ or equivalent.  

| **Partially Correct Response** |  
|--------------------------------|---|
| **10**                         | Expression for red tiles correct as in 20 but **not** expression for total tiles  
| **11**                         | Expression for total tiles correct as in 20 but **not** expression for red tiles  

| **Incorrect Response** |  
|------------------------|---|
| **70**                 | Incorrect expression including $n$ for red tiles or total or both (includes incorrect attempts to express red tiles as difference from total tiles)  
| **79**                 | Other incorrect (including crossed out, erased, stray marks, illegible, or off task)  

| **Nonresponse** |  
|----------------|---|
| **99**         | Blank  

- **All codes considered as different response categories.**
Regularization path of item M032761

non-adaptive

adaptive

slope intercept
incorrect partially correct correct

Michela Battauz
Regularized estimation of the nominal model
July 10, 2019 11 / 20
Simulation studies

- 24 parameters
- $\alpha_{12} = \alpha_{13}$
- $n =$ 200, 500, 1000, 5000
- 500 replications
Results of the simulation study

Number of cases out of 500 in which $\alpha_{12}$ and $\alpha_{13}$ are fused at the selected value of $\lambda$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>200</th>
<th>500</th>
<th>1000</th>
<th>5000</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-adaptive</td>
<td>53</td>
<td>43</td>
<td>28</td>
<td>19</td>
</tr>
<tr>
<td>adaptive</td>
<td>264</td>
<td>287</td>
<td>336</td>
<td>357</td>
</tr>
</tbody>
</table>
Absolute bias of penalized estimates versus MLE

- **n = 200**
  - MLE vs. penalized estimates
  - △ intercept
  - ○ slope

- **n = 500**
  - MLE vs. penalized estimates
  - △ intercept
  - ○ slope

- **n = 1000**
  - MLE vs. penalized estimates
  - △ intercept
  - ○ slope

- **n = 5000**
  - MLE vs. penalized estimates
  - △ intercept
  - ○ slope

**Intercepts:**
- Type: Regularized estimation of the nominal model
- Author: Michela Battauz
- Date: July 10, 2019
- Page: 14/20
Absolute bias of adaptive penalized estimates versus MLE

- n = 200
- n = 500
- n = 1000
- n = 5000

\[ \text{MLE} \quad \text{penalized} \]

△ intercept  ○ slope
Root mean square error of penalized estimates versus MLE

\[ \text{MLE} \quad \text{penalized} \]

\( n = 200 \)

\( n = 500 \)

\( n = 1000 \)

\( n = 5000 \)

\[ \triangle \text{ intercept} \quad \circ \text{ slope} \]
Root mean square error of adaptive penalized estimates versus MLE

- $n = 200$
- $n = 500$
- $n = 1000$
- $n = 5000$

$\triangle$ intercept
○ slope
R package regIRT

- Available at https://github.com/micbtz/regIRT.
- Currently implements the regularized nominal model.
- The main function is nominalmod
  - non-adaptive penalization
    
    ```r
    > mod_nonadp <- nominalmod(data = nomdata, D = 1,
    + parini = par, lambda = seq(0, 3, length = 30),
    + pen = "lasso", adaptive = FALSE)
    ```
  - adaptive penalization
    
    ```r
    > mod_adp <- nominalmod(data = nomdata, D = 1,
    + parini = par, lambda = seq(0, 3, length = 30),
    + pen = "lasso", adaptive = TRUE, parW = parMLE)
    ```

- Function nominalCV performs cross-validation, function regPath plots the regularization path.
The proposal

- can be used to **collapse** response categories,
- provides **regularized estimates** of **slopes** and **intercepts**, 
- reduces **bias** in small samples,
- improves **efficiency**.

- Currently working on the extension to the **multidimensional** nominal model.


Thank you for your attention!